Applied Machine Learning - Intro

Annalisa Marsico
OWL RNA Bioinformatics, MPI Molgen Berlin
Freie Universität Berlin
20.04.16

Who am I?

"Laurea" in Physics (5 years)

Thesis: „From conventional tests to adaptive tests: application of artificial neural networks to ability evaluation"

Master in Bioinformatics (1.5 years)
PPNEMA: database for sorting and classificaiton of rRNA genes in nematodes

PhD in Bioinformatics (4 years)
Prof. Michael Schroeder (TU Dresden) & Prof. Marino Zerial (MPI MGC)

- **clustering algorithm** for de novo motif discovery
- MeMotif: database of sequence / structure motifs
- pattern classification of unolding pathways from biophysical data

Post-doc at the Max Planck for molecular Genetics Berlin (3 years)
Prof. Martin Vingron

- PROmiRNA: semi-supervised learning of miRNA promoters
- linear model for miRNA processing from sequence signal
- RNA-seq and ChIP-seq data for infection processes (SFB TR48)

The RNA Bioinformatics group

- Since **July 2014** Junior Professor for High-Throughput Genomics (FU) & group leader of the RNA Bioinformatics group at MPI Molgen

http://www.molgen.mpg.de/2733742/RNA-Bioinformatics

Additional co-lecturers

Stefan Budach, PhD student

- Bachelor & Master in Bioinformatics, Freie Universität Berlin
- Bachelor thesis in Martin Vingron‘s lab on statistical modeling of batch effects in RNA-Seq data
- Student assistant in my lab for a year (2014) predictive model of miRNA-eQTLs
- Master thesis in Knut Reinert‘s lab on parallelization in SeqAn
- Since November 2015, IMPRS PhD student in the RNA Bioinformatics Lab

Wolfgang Kopp, PhD student

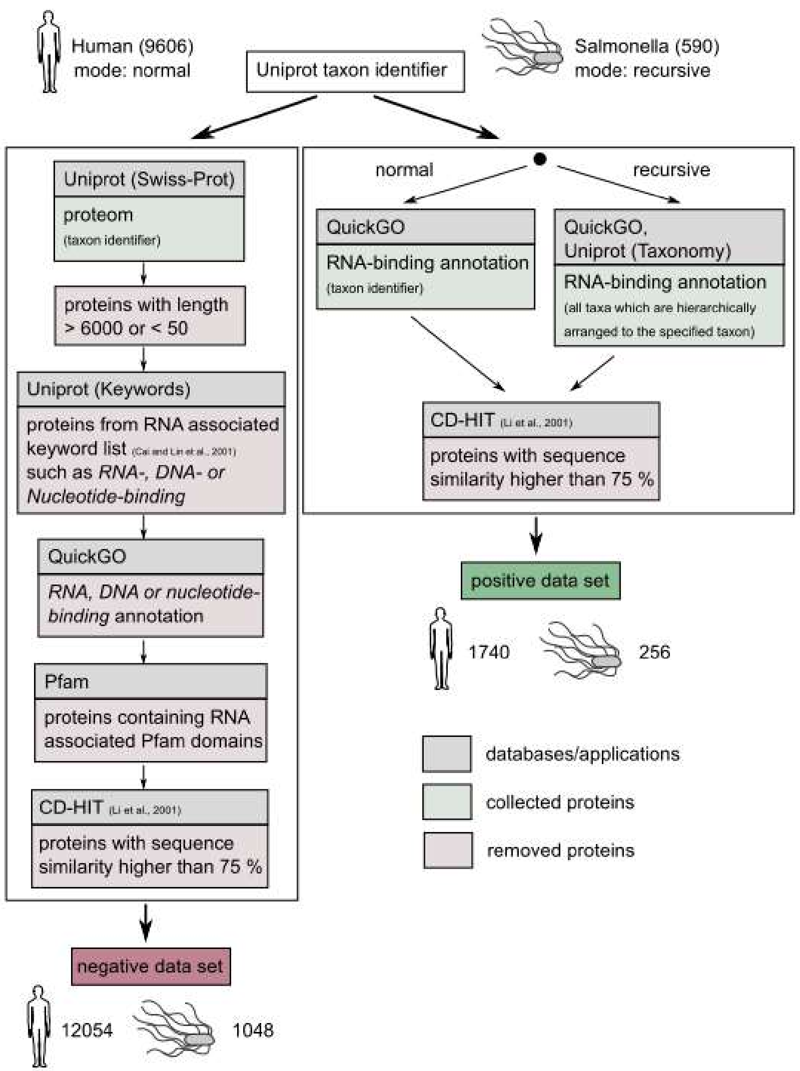
- Bachelor & Master in Biomedical Engineering, Technical University of Graz (Austria)
- student assistant for the ‚Computational Intelligence‘ course
- Since 2012 PhD student of the IMPRS in the lab of Martin Vingron (MPI Molgen) Projects: statistical models for pattern occurrences in DNA sequences, ChIP-seq time-series analysis

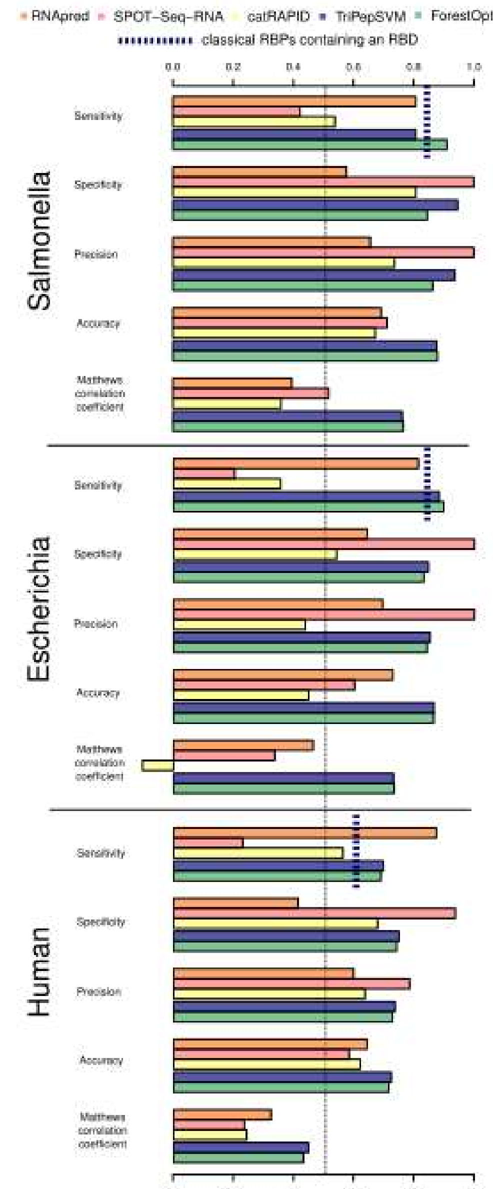
Work in our Lab

- RNA Binding Proteins (RBPs) prediction
- RBPs target identification
- Target specificity (de novo motif discovery)

RBPs prediction

TripSVM - A tool for species-specific RBP prediction

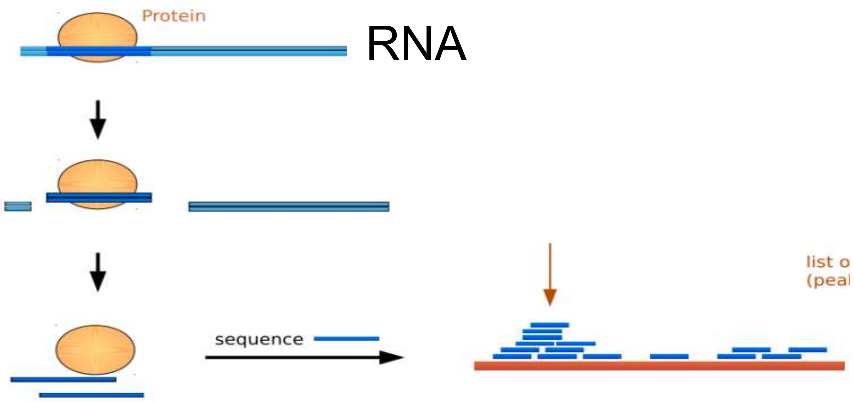


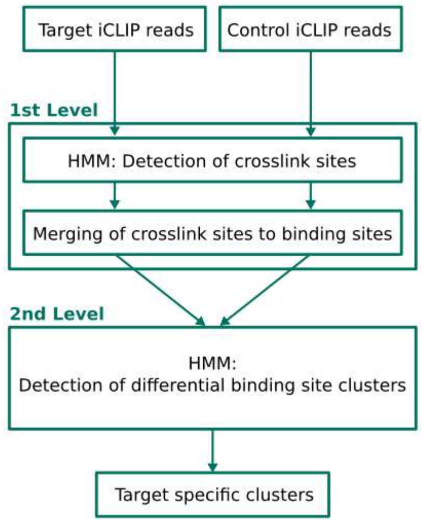


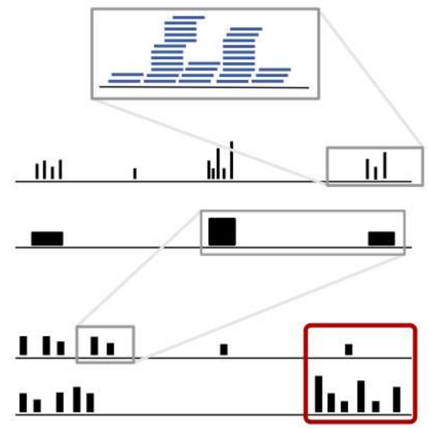
Annkatrin Bressin, Benedikt Beckmann

RBPs and their targets

Capturing target-specific, clustered protein-RNA interaction footprints from iCLIP-seq



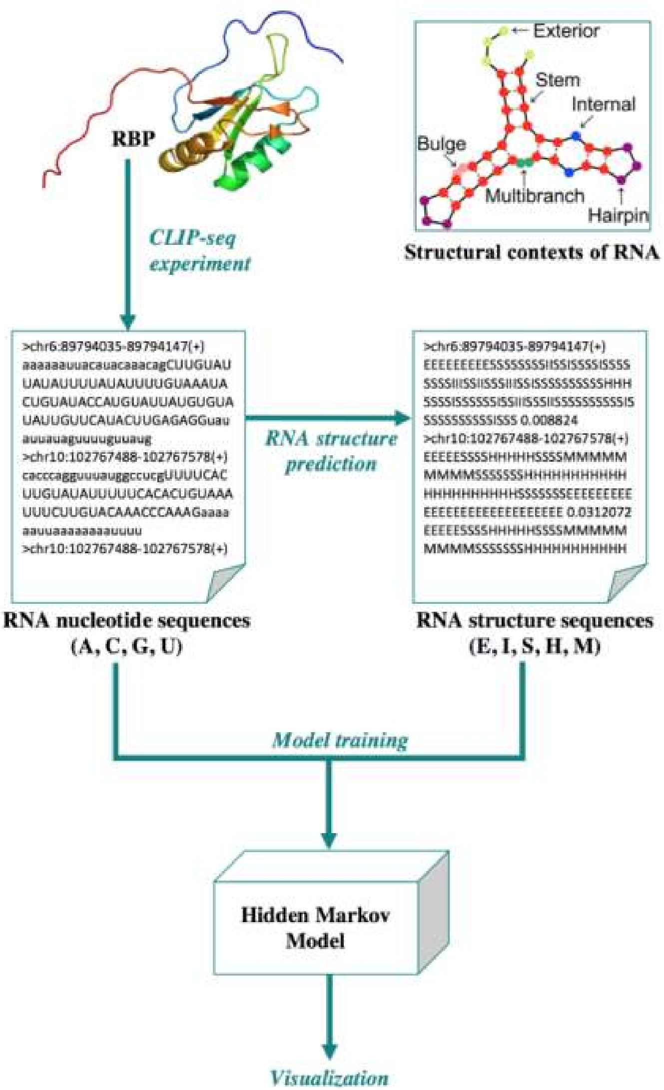




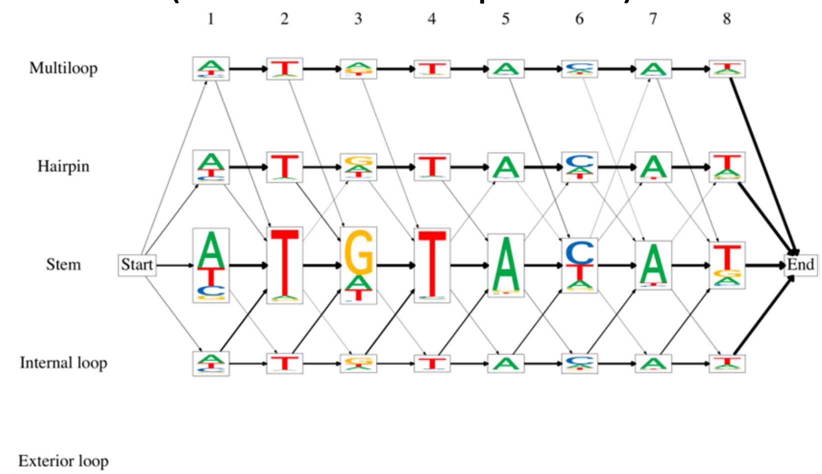
Sabrina Krakau, Hugues Richard (PMC Paris)

RBPs and their targets

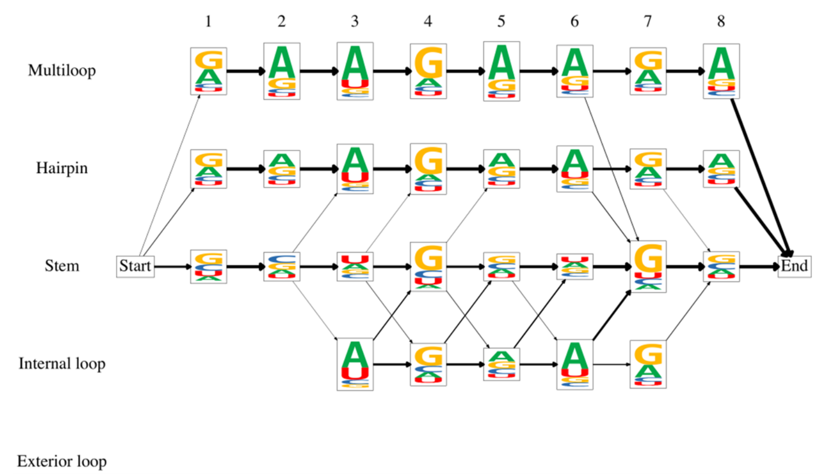
A tool for de novo sequence-structure motif discovery



PUM2 (translational repressor) model



SFRS1 (splicing factor) model

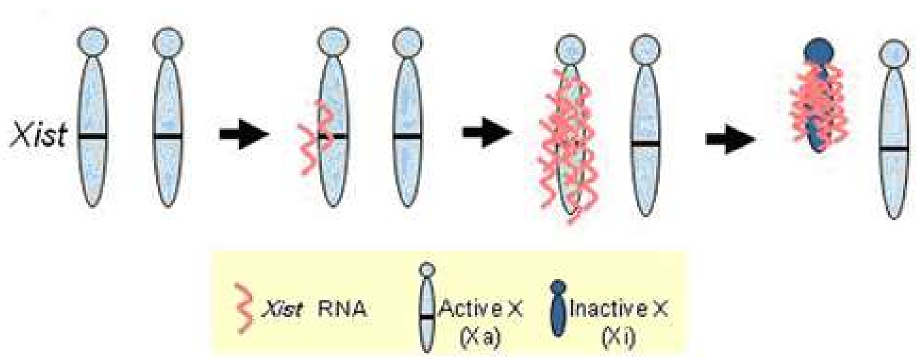


David Heller, Ralf Krestel (Uni Potsdam), Uwe Ohler (MDC)

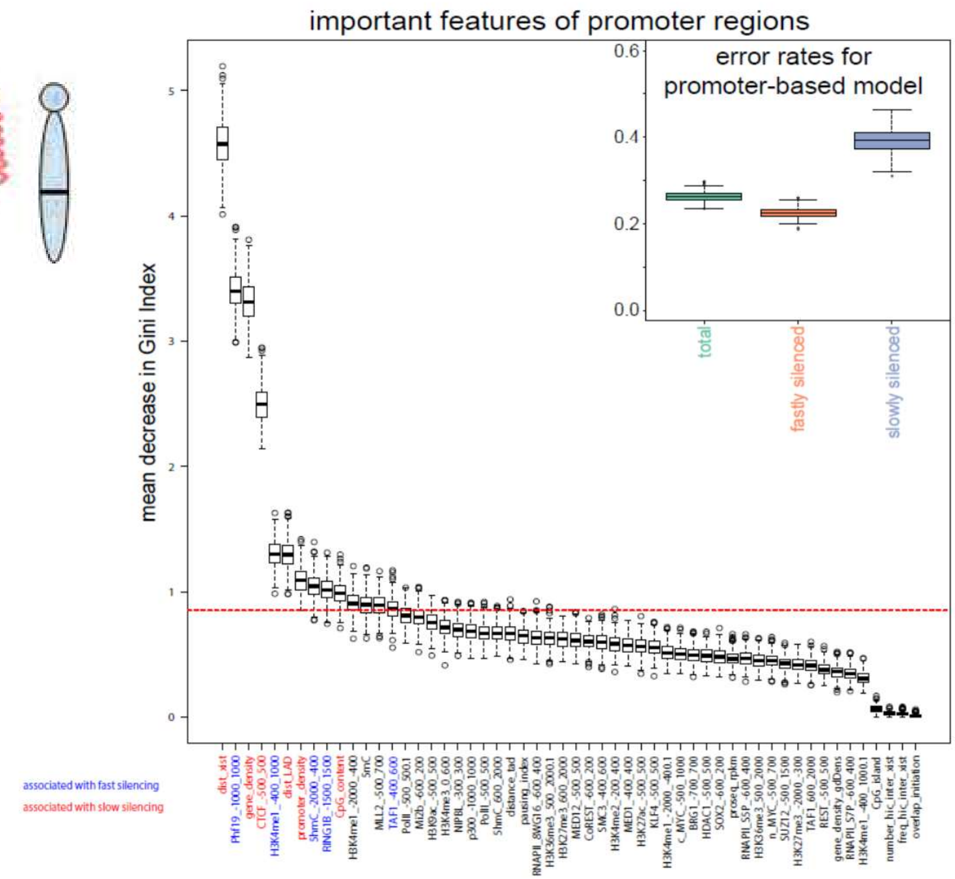
Work in our Lab

- Regulation and function of lncRNAs
- microRNA-eQTL prediction

Predicting X-chromosome silencing kinetics



Random Forest prediction



Lisa Barros, Edda Schulz

Deep Learning

Predicting DNAse I Hypersensitivity Sites from Dna Sequences Using Deep Learning Techniques
- Convolutional Boltzmann Machines

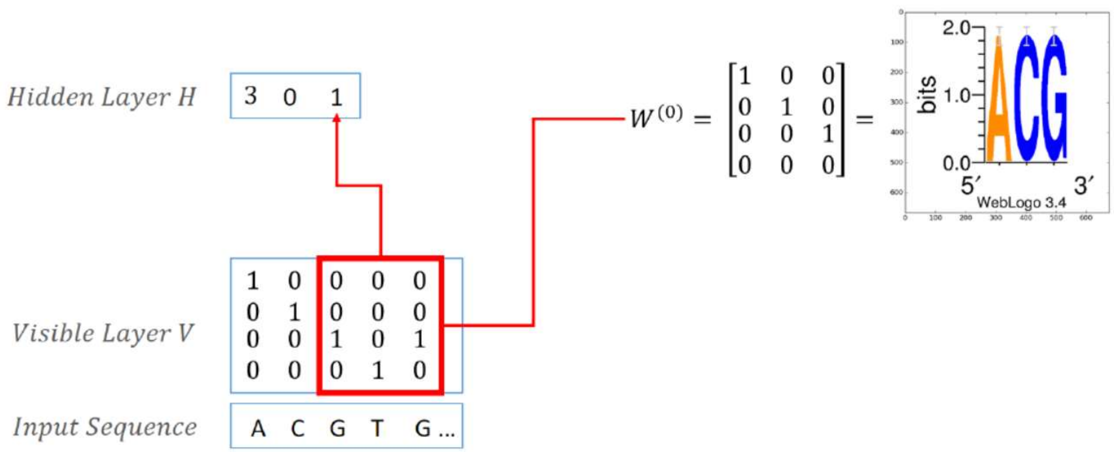


Figure 3: Convolutional layer of a cRBM. The convolution produces a smaller hidden layer H that contains the motif hits.

Roman Schulte-Sasse, Wolfgang Kopp

The Plan

Lecture	Application	Method
8 June	Cancer classification from gene expression	Partial Least Square Regression (PLSR)
15 June	Cancer classification from clinical data / DNA methylation	SVMs
22 June	RBP binding site prediction from RNA sequences	String kernel SVMs
29 June	Promoter prediction from epigenetic features	Semi-supervised learning: co-training algorithm
06 July	Cancer location prediction from microRNAs	Neural Networks (multi-class classification)
13 July	Applications of DP in Bioinformatics	Concepts of deep Learning
20 July	Discuss applications of DP in Bioinformatics	Discuss applications of DP in Bioinformatics

Deep Learning Papers

1. Deep Architectures for Protein Contact Map Prediction
   http://www.ncbi.nlm.nih.gov/pubmed/22847931

2. Toxicity Prediction Using Deep Learning
   http://arxiv.org/pdf/1503.01445.pdf

3. Deep Learning of tissue-regulated splicing code
   http://www.ncbi.nlm.nih.gov/pubmed/24931975

4. Predicting effects of non-coding variants with deep-learning-based sequence model
   http://www.ncbi.nlm.nih.gov/pubmed/26301843

5. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning
   http://www.ncbi.nlm.nih.gov/pubmed/26213851

Link to course material

http://www.molgen.mpg.de/3434810/Applied-Machine-Learning

Contacts:

marsico@molgen.mpg.de
budach@molgen.mpg.de
kopp@molgen.mpg.de